

DATABASE MIGRATION ANALYSIS FOR DATA ASSURANCE USING D2LOADER

ALVIN JASON A. VIRATA, JONATHAN M. CABALLERO & DAWN IRIS CALIBO

College of Information Technology Education, Technological Institute of the Philippines, Philippines

ABSTRACT

Data migration process can take place in different forms such as moving data, application or business elements from an organization's onsite computer to the cloud, vice-versa or between two cloud environments. There are lots of issues that can be identified such as security, different standards, compatibility of computing resources, performance, formats. The primary objective of the study is to analyze data migration towards data assurance using D2Loader, to adopt compatibility of the new platform. Specifically, it will identify the pattern of queries required for data migration; validate the migrated data effectively; and ensure the synchronize pattern of data used to test data reliability and integrity for data assurance.

KEYWORDS: *Data Migration, System Transition, Intensive Query*

Received: Jun 22, 2017; **Accepted:** Jul 07, 2017; **Published:** Jul 12, 2017; **Paper Id:** IJCSEITRAUG20173

INTRODUCTION

Background/ Objectives and Goals

Data migration is considered a complex process that requires a robust methodology. In the same manner, databases form a major part of the infrastructure that underpins different storage for all data, so that, it can be accessed through the application. Database migrations are activities that need to be visualized, in order to test the data integrity and reliability to ensure that no data will be lost.

As defined by Vaidya and Pundkar, data migration is the process of transporting the data from one place to another place. Data migration process can take place in different forms such as moving data, application or business elements from an organization's onsite computer to the cloud, vice-versa or between two cloud environments. There are lots of issues that can be identified such as security, different standards, compatibility of computing resources, performance, formats etc.

Further, dealing with data migration will require a well-managed and robust approach to minimize the risk. Considering that an object with no well-defined relationship to the object, will be considered as not valuable (Liang et al, 2015). Since it requires a clear understanding as to how the recovery options, migration methodology for each stage will be applied. Identifying the appropriate and effective data migration and integration processes will produce a thorough landscape analysis, to understand how each system and subsystem works, and how data within each system is being structured.

As observed, there were issues that were noted and required attention and a particular solution. Identifying the existing platform framework, that supports the existing system, was the primary consideration. It was found out that the current platform framework used was Microsoft 2.0 version, while new version has been released already by Microsoft, which is 7.0. The need for upgrade and improvement was the problematic part, since the existing

system has compatibility issues and software development support has already reached its end. To resolve the issue, it was mandatory that an application be developed for database migration, which is the D2Loader. D2Loader is a third party system that served as the bridge to transfer the old data records to the new database design to ensure secured migration of data.

The primary objective of the study is to analyze data migration towards data assurance, using D2Loader to adopt compatibility of the new platform. Specifically, it will identify the pattern of queries required for data migration; validate the migrated data effectively; and ensure the synchronize pattern of data used to test data reliability and integrity for data assurance.

Extract, Transform and Load (ETL) is a well-known term in data warehousing (Poe et al.). In figure 1 below, it describe the conceptual working of ETL. ETL method is widely applicable since it represents a schema to schema mapping, purely from the data-oriented view. It is not necessary to know the inner working of the source and target system/application that operates on the data (Hsieh et al.)

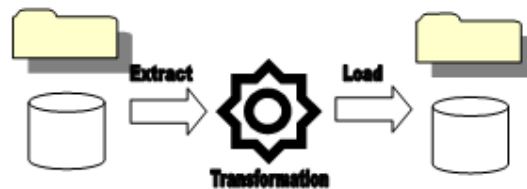


Figure 1: Conceptual Working of ETL

Identifying the architecture in data migration is one of the methods to be considered. It works on block level data migration, which retained data integrity; rapid migration between storage devices with no application downtime is done with maximum efficiency and effectiveness. Defining and measuring the quality of data is very complex. According to Vaidya and Punkdar, in data migration, there are lots of issues that must be considered such as security, different standards, compatibility of the computing resource, formats, performance, cost, etc. In this note, to identify the quality of the data extracted, loaded and transformed the following characteristics are considered (1) Accuracy (2) Completeness (3) Consistency (4) Availability and (5) Traceability.

METHODS

Iterative model is extremely easy to utilize. It takes basic usage of a little arrangement of the framework prerequisites, and after that intuitively improves the developing modules, until the real framework is actualized and prepared to be conveyed. The life cycle of this modeling approach does not focus to the detail with full specification of prerequisite. In the modeling approach, primary concentration is toward the improvement of output first, by determining and executing simply little piece of the product(output), which can be consistently looked into with a specific end-goal, to distinguish necessary requirements for the improvement, usefulness, and functionality of the system.

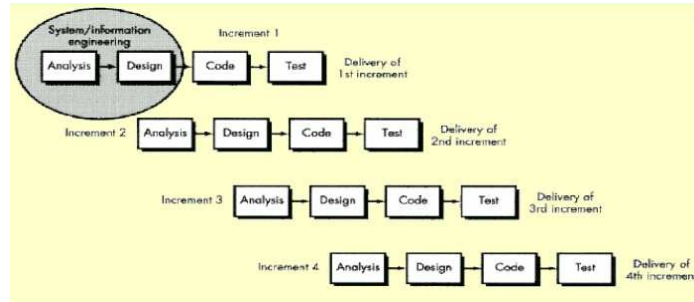


Figure 2: Database Migration Development Life Cycle by M. Heusser and T. Salt

Feedback mechanism is the core focus, when software development is concerned. Feedback result helps the programmer and system analyst identify the changes and improvement requirement. Thus, the analysis and design phase are working together to clearly define the information requirement, knowledge requirements and the re-engineering requirement of the system.

Analysis: The researcher has come to decide to identify the functional and non-functional requirements before the development of Database Migration System, to ensure completeness and accuracy of the data to be transferred to the new database.

Design: The design interface was carefully checked to ensure that there are no data left behind before the migration process.

Figure 3: Database Migration Interface (D2Loader)

Implement: After the design, the database migration module was implemented.

Test: When the code was already written, there are four stages of testing that was done: (1) Alpha Testing (2) Beta Testing (3) Usability Testing and (4) User Acceptability Testing.

Install: Once Database Migration is fully tested for its functionality, it was immediately executed based on the deployment plan, as defined in the requirement phase, therefore, a bulk database migration was done.

Maintain: After installment, maintenance is conducted to ensure that the system will be functional.

Application Module Scope

- **Upload:** In this module, the user can upload any type of files as swift storage supports, both structured and unstructured data. Like Google drive account, user can upload a file and send it to other users.

- **Transfer/Migrate:** Once the file gets uploaded, the file is migrated to the destination swift storage server demand, when the user commands.

Database Migration Process

Entry of Student # (e.g.: 20121000)

Step 1: Recovery of student info (Name, Address, Contact Info, Family [Father, Mother], Last School, etc.)

Step 2: Recovery of Item Payments (Tuition Fees, Miscellaneous fees, Project Fees, etc.) per Sem. & School Yr

Step 3: Recovery of Advised Subjects per Semester & School Year

Step 4: Recovery of Credited Subjects (Subjects from other school [Transferee])

Step 5: Recovery of Enrolled Subjects with Grades per Semester & School Year

Step 6: Recovery of Completion Subjects (Retake Subjects)

Step 7: Recovery of Transcript Logs (TOR release Logs)

Step 8: Recovery Successful, Uploading Process Ready.

Step 9: Uploading of Data (Student Info. Payment Logs, Advised & Enrolled Subj w/ Grades per Sem & S. Y.)

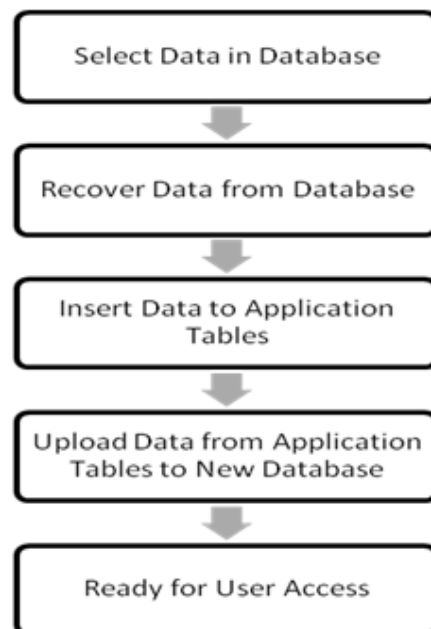
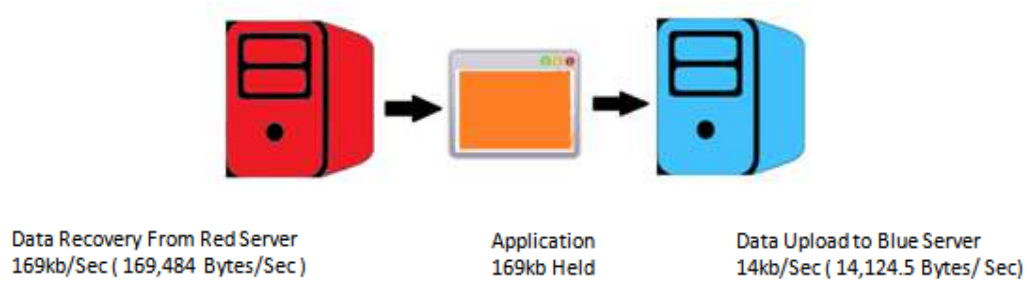


Figure 4: Database Migration Design

RESULTS

Data Migration (Initial Testing) Implementation

This step was used to determine, if the individual migration is successful. In this process, it will also validate the completeness and accuracy of the data being migrated from the old database to the new database

**Figure 5: Data Migration Physical Model****Table 1: Shows the Result of Data Migration Process from Existing Database to New Database**

Data Size	Recovery Time	Upload Time	Total Migration Time
1 Byte	0.000169ms	0.000014ms	0.000173ms
1 KB	0.169ms	0.014ms	0.173ms
1 MB	169ms	14ms	173ms

After a series of testing conducted, the speed of transferring data to another database is not a hassle. Since the design interface was tailored-fit to the newly developed system

Analysis of Database Migration Transition

Aside from identifying the speed of transferring database record to the new database design, the following issues or concerns were resolved:

- The data types should be assessed and checked to ensure the completeness of the data migration. Applying data type conversion should be put into consideration to avoid data loss.
- Special characters involved in the data records are also one of the areas to be validated in the record. Creating a function that will detect the special characters should be reinforced to convert data record and meet SQL compatibility.
- Migration process should not be done in a single entry method, but a bulk data migration must be implemented. Bulk data migration should undergo series of testing to ensure the database migration process.
- An intensive query database management was implemented, and parallel transferring was applied for both database server to compliment the completeness and accuracy of the data record for the new system database schema.

Table 2: Sql Query Method

SQL Method	Action	Notes
SELECT	Read	-extracts data from the database (e.g. SELECT Name FROM Customers)
INSERT INTO	Create	-Create record inside table in database (e.g. INSERT INTO Customer VALUES (Juan, Dela Cruz)
WHERE	Filter	-Filter information by Desired Output (e.g. SELECT Name FROM Customers WHERE Gender = Male)

Proper filtering and query are necessary to ensure that data integrity was observed. The accuracy of data was one of the areas, which were observed to make sure the completeness of the data being transferred to the new database. Also, some data types were converted and redesign to compliment with the new database design.

Database migration is considered a very crucial process that requires a diligent handling of data records. Data records are complicated, especially when a special character is involved in the data such as the single or double apostrophe (quotation), which is one of the limitations of the SQL query method. The date and time format are also areas that must be considered to ensure integrity of the data being transferred. At the same time, during database migration, it was noted that transformation of character set should be considered to avoid data lost. A function that will detect special character is necessary to prevent error during query method.

ACKNOWLEDGMENTS AND LEGAL RESPONSIBILITY

The researchers would like to acknowledge the professors and colleagues from College of Information Technology Education for giving their time and effort to make this paper possible. Above all, we would like to give ALL the glory and honor to the Almighty God.

REFERENCES

1. Calil, A., & dos Santos Mello, R. (2012, September). *Simple SQL: a relational layer for Simple DB*. In East
2. *European Conference on Advances in Databases and Information Systems* (pp. 99-110). Springer Berlin Heidelberg.
3. Hsieh, C. Y., Hsieh, H. A., & Cheng, Y. C. (2016, May). *A method for web application data migration based on*
4. *RESTful API: A case study of ezScrum*. In *Applied System Innovation (ICASI), 2016 International Conference on* (pp. 1-4). IEEE.
5. Liang, D., Lin, Y., & Ding, G. (2015, December). *Mid-model Design Used in Model Transition and Data*
6. *Migration between Relational Databases and NoSQL Databases*. In *Smart City/SocialCom/SustainCom (SmartCity), 2015 IEEE International Conference on* (pp. 866-869). IEEE.
7. Heusser, M., & Sait, T. (2015). *General Idea of Iterative Models- Spiral Model*. Retrieved 2017, from Scribd:
8. <https://www.scribd.com/presentation/72974150/Final-Spiral-Model-97>
9. Verma, E. (2015, May). *SDLC Iterative Model Phase, Application, Advantages and Disadvantages*. Retrieved
10. 2017, from yuvayana: <http://er.yuvayana.org/sdlc-iterative-model-design-phase-applications-advantages-and-disadvantages/>
11. Wang, Q., Liu, N., & Yu, C. (2012, August). *Research and Practice of university database Migration*.
12. In *Information Technology in Medicine and Education (ITME), 2012 International Symposium on* (Vol. 1, pp. 31-34). IEEE.
13. Vaidya, R., & Pundkar, S. (2015, September). *Optimized data migration within OpenStack cloud*. In *Bombay*
14. *Section Symposium (IBSS), 2015 IEEE* (pp. 1-5). IEEE.
15. Shin, S. K., & Sanders, G. L. (2006). *Denormalization strategies for data retrieval from data warehouses*. *Decision Support Systems*, 42(1), 267-282.